

February 24, 2020

**Asians do not exhibit elevated expression or unique genetic polymorphisms for ACE2, the cell-entry receptor of SARS-CoV-2**

Ying Chen<sup>1,2</sup>, Kejia Shan<sup>1,2,3</sup>, and Wenfeng Qian<sup>1,2,3,\*</sup>

<sup>1</sup> State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

<sup>2</sup> Key Laboratory of Genetic Network Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China

<sup>3</sup> University of Chinese Academy of Sciences, Beijing 100049, China

\* Correspondence to:

Wenfeng Qian

Institute of Genetics and Developmental Biology

Chinese Academy of Sciences

Beijing 100101, China

Email: [wfqian@genetics.ac.cn](mailto:wfqian@genetics.ac.cn)

## ABSTRACT

The recurrent coronavirus outbreaks in China (SARS-CoV and its relative, SARS-CoV-2) have raised speculations that perhaps Asians are somehow more susceptible to these coronaviruses. Here, we test this possibility based on an analysis of the lung-specific expression of *ACE2*, which encodes the known cell-entry receptor of both SARS-CoV and SARS-CoV-2. We show that *ACE2* expression is not affected during tumorigenesis, supporting that the abundant transcriptomes in cancer genomic studies can be informatively used to study *ACE2* expression among diverse individuals without cancer. We find that *ACE2* expression in the lung increases with age, but is not associated with sex. Further, Asians do not differ from other populations for *ACE2* expression and do not harbor unique genetic polymorphisms in the *ACE2* locus. Thus, beyond illustrating an innovative method for assessing the potential impacts of demographic factors for non-cancer diseases from large-scale cancer sample datasets, our statistically robust findings emphasize that individuals of all races require the same level of personal protection against SARS-CoV-2.

## KEYWORDS

SARS-CoV-2; cell-entry receptor; *ACE2*; The Cancer Genome Atlas; susceptibility; demographic factors; polymorphism

## INTRODUCTION

The outbreak of coronavirus disease 2019, which is caused by SARS-CoV-2, has led to significant illness and death and has been designated a global health emergency by the World Health Organization. It is clear that SARS-CoV-2 is a close relative of SARS-CoV (Xu et al., 2020), which caused the well-studied severe acute respiratory syndrome in 2003. However, any demographic factors which may predict differential susceptibility to these coronaviruses remain poorly understood. In particular, since both SARS-CoV and SARS-CoV-2 epidemics broke out in China, it has been speculated that East Asians may be relatively more susceptible to these coronaviruses (Zhao et al., 2020).

Biologically, research from around the time of the SARS epidemic in 2003 revealed that SARS-CoV enters cells through a protein known as Angiotensin-converting enzyme 2 (ACE2) (Kuba et al., 2005; Li et al., 2003), whose native function is to play a role in the renin-angiotensin system regulating blood pressure and so forth (Donoghue et al., 2000). Susceptibility to SARS-CoV was positively associated with the *ACE2* expression level among cells in the lung, as well as among nine diverse cell lines (Hofmann et al., 2004; Jia et al., 2005). Another study showed that, on the one hand, overexpressing *ACE2* in cell lines promoted efficient replication of SARS-CoV, while on the other hand use of neutralizing antibodies against ACE2 inhibited viral replication, and did so in a dose-dependent manner (Li et al., 2003). In humans, profiling revealed *ACE2* expression in alveolar epithelial cells (Hamming et al., 2004), which are understood as the primary site of SARS-CoV infection in the lung (Kuiken et al., 2003). And the study using an *ACE2* knockout mouse model showed that mice lacking functional ACE2 exhibited reduced SARS-CoV levels in the lung (Kuba et al., 2005).

Very recent work has indicated that ACE2 is also used by SARS-CoV-2 to enter cells, and structural analyses revealed apparently strong binding affinity between the ACE2 protein and the spike protein of SARS-CoV-2 (Wrapp et al., 2020; Xu et al.,

2020; Yan et al., 2020). Furthermore, it was demonstrated that HeLa cells were only susceptible to infection by SARS-CoV-2 when *ACE2* was expressed (Zhou et al., 2020). Thus, *ACE2* expression in lung cells is understood as by far one of the most promising indicators for susceptibility to infection by SARS-CoV-2.

An analysis using the data of the genotype-tissue expression (GTEx) project failed to identify any quantitative trait loci associated with *ACE2* expression in the lung (<https://gtexportal.org/home/gene/ACE2>); Asian-biased expression for any genes was not detected (Mele et al., 2015), potentially owing to the small number of samples from Asians (1.3% of all samples) in the GTEx project. In contrast, we reasoned that data for thousands of lung samples from a great diversity of individuals of various ages, sexes, and races are present in the database of The Cancer Genome Atlas (TCGA) and other cancer genomics studies. Pursuing this, we here investigated potential demographic factors affecting *ACE2* expression based on cancer transcriptome data. We show that *ACE2* expression is not affected during tumorigenesis, and find that Asians do not have higher *ACE2* expression than other populations in a number of organs, including the lung. Furthermore, we did not detect any genetic polymorphisms in the *ACE2* locus specifically enriched amongst Asians. Thus, beyond helping to better understand the basic biology of coronavirus infection, our study can inform global efforts to develop efficient protection strategies against SARS-CoV-2.

## RESULTS AND DISCUSSION

TCGA includes data for lung transcriptomes of two cancer types: lung adenocarcinoma (LUAD) and lung squamous cell carcinoma. We focused on the LUAD samples in this study because LUAD is known to be more likely developed from alveolar cells (Lin et al., 2012; Xu et al., 2012). The expression level of *ACE2* was similar between primary LUAD tumors and adjacent normal tissues (**Figure S1A–B**) and was not associated with survival probability or pathologic stage (**Figure**

**S1C–D**), indicating that *ACE2* expression is not apparently affected during tumorigenesis.

Pursuing this further with additional cancer types, we determined that the expression level of *ACE2* in primary tumor samples did genuinely reflected the levels of adjacent normal tissues across 19 cancer types in the TCGA (Pearson's correlation coefficient  $R = 0.8$ ,  $P = 3 \times 10^{-5}$ , **Figure S1E** and **Table S1**). These findings support that the LUAD transcriptome data can be used to study *ACE2* expression among cancer-free individuals. However, there were only seven Asian LUAD samples in the TCGA; to address this paucity of relevant samples, we expanded the sample size for Asians by including 260 Chinese LUAD transcriptomes (Chen et al., 2020a).

To determine if cancer transcriptome data can detect differential gene expression with sufficient sensitivity among cancer-free individuals, we used the 38 genes previously reported from the GTEx project that showed sex- or race-specific expression in the lung (Mele et al., 2015) as the gold standard. Among them, 37 genes are expressed in the LUAD samples, and all of these (37 out of 37) showed the same trends for significant differential expression as reported using the GTEx data (**Figure S2**), findings highlighting that our approach is highly sensitive for detecting demographic factors affecting gene expression in lungs.

Analysis in LUAD samples revealed that *ACE2* expression was positively correlated with age among middle-aged and older adults (Pearson's correlation coefficient  $R = 0.13$ ,  $P = 2 \times 10^{-4}$ ,  $N = 835$ , **Figure 1A**), increasing by  $\sim 1.2$  times with every 10-year increase in age. This result may help partially explain the observation that the elderly are more susceptible to SARS-CoV-2 (Chen et al., 2020b; Huang et al., 2020; Li et al., 2020; The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team, 2020). Sex was not associated with *ACE2* expression ( $P = 0.24$ , the two-tailed Mann-Whitney  $U$  test, **Figure 1B**), echoing reported epidemiological findings that both sexes are equally likely to be infected by SARS-CoV-2 among 72,314 patients (The Novel Coronavirus Pneumonia Emergency Response

Epidemiology Team, 2020).

The average expression level of *ACE2* among Asians was not significantly different from that among individuals of African or European ancestry (hereafter referred to as African or European,  $P = 0.42$  and  $0.96$ , respectively, assessed using two-tailed Mann-Whitney  $U$  tests, **Figure 1C**). Nor was any race-specific difference detected by a linear model which predicted *ACE2* expression levels based on multiple demographic factors including age, sex, and race (**Table S2**); the variance in the log<sub>2</sub>-transformed *ACE2* expression level among all LUAD samples was 3.53 (i.e., standard deviation = 1.88), and race explained only ~ 0.1% of this variance ( $P = 0.64$ , the analysis of variance). These findings and data-driven predictions from our study refute a previously reported conclusion about the biased *ACE2* expression that was based on a single Asian sample (Zhao et al., 2020).

There are some caveats to bear in mind with the present study. First, we focused on *ACE2* expression in the lung, but *ACE2* is expressed in other tissues/organs as well (**Figure S3**), with especially strong expression in the gastrointestinal tract and kidney. These tissues/organs were also reported to be infected by SARS-CoV (Gu et al., 2005), and SARS-CoV-2 has been detected in patients' stool (Holshue et al., 2020). To address this, we further examined the potential Asian-biased expression of *ACE2* in other cancer types. No such Asian-bias was detected for any cancer types with higher *ACE2* expression levels than LUAD, including stomach, colon, and kidney cancers (**Figure 2**). Nevertheless, we cannot rule out the possibility that maybe *ACE2* expresses at higher levels in Asians in some other as-yet-unexamined tissues.

A second caveat relates to our exclusive focus at the gene expression level; it is possible that some Asian-specific genetic variations in the coding sequence of *ACE2* may also affect the cell-entry efficiency of viruses. To explore the potential impact of population-specific genetic variations, we retrieved all of the genetic variation data available for the *ACE2* locus from the 1000 Genomes Project (The 1000 Genomes Project Consortium, 2015), which include samples from Africans, (native) Americans,

East Asians, Europeans, and South Asians. We focused on the polymorphisms in East Asians, since both coronavirus epidemics broke out in East Asia. All 22 of the missense or stop-gained variations were present at a low frequency among East Asians (< 1%, **Figure 3**). Other variants ( $N = 690$ , **Table S3**) are mostly located in introns, and the frequencies of these variants among East Asians were highly correlated with those among all populations (Pearson's correlation coefficient  $R = 0.996$ ,  $P < 2.2 \times 10^{-16}$ , **Figure 3**). For instance, we found that among the 37 detected polymorphisms having an alternate allele frequency > 10% in East Asians, none (0 out of 37) had reached a frequency two times higher than all populations (**Figure 3**). We reached a similar conclusion about the lack of any East-Asian-specific enrichment for *ACE2* genetic variants based on analyzing data from the Genome Aggregation Database (gnomAD, **Figure S4**) (Karczewski et al., 2019).

Another possibility is that other layers of regulation (e.g., translational efficiency, protein modification, folding, or subcellular localization) exist in addition to regulation at the mRNA level. While *ACE2* protein levels are largely correlated with mRNA levels among tissues (<https://www.proteinatlas.org/ENSG00000130234-ACE2/tissue>) (Uhlen et al., 2015), whether these *ACE2* proteins are located on the cell membrane remains unclear. Therefore, future investigations should examine potential race-specific bias for the accumulation of functional *ACE2* on the plasma membrane. It would also be informative to investigate if co-receptors of SARS-CoV-2 exist and, if so, to determine if their abundances vary among populations. The identification of host cell features that affect the efficiency of viral amplification will also help improve accuracy for predicting susceptibility to coronaviruses in the future.

Based on the available data, we conclude that Asians do not express *ACE2* at a higher level than other populations and do not bear unique genetic polymorphisms in *ACE2*. The recurrent coronavirus outbreaks in China may be better explained by the high diversity of coronaviruses and their animal hosts, or perhaps by Chinese food culture (Fan et al., 2019). Our study, therefore, cautions against any use of race to predict susceptibility to SARS-CoV-2 among individuals; that is, individuals of all

ances require the same level of personal protection against SARS-CoV-2.

## METHODS

All available open RNA-seq data and clinical data for LUAD samples in TCGA were retrieved from <https://www.cancer.gov/tcga>. The RNA-seq data were obtained in the FPKM-UQ (fragments per kilobase of transcript per million mapped reads of the upper quartile gene) format, which is known to perform better in cross-sample comparisons and differential expression analyses. Transcriptomes of additional East Asian samples (Chen et al., 2020a) were retrieved from OncoSG (<https://src.gisapps.org/OncoSG/>) under the dataset “Lung Adenocarcinoma” (GIS, 2019); FPKM-UQ values for these samples were calculated from the numbers of reads mapped to individual genes. Notably, the FPKM-UQ values of expressed genes were globally higher in the data from OncoSG than from the TCGA data (**Figure S5A**); therefore, median normalization was further performed to help ensure comparability of the expression levels from these two data sources (**Figure S5B**). The annotation of demographic information for each sample and statistical analyses were performed using in-house R scripts. TCGA contains seven Asian LUAD samples and does not further distinguish East Asians from South Asians in its annotation. Nevertheless, after including the 260 Chinese LUAD samples, the majority (> 97%) of Asian samples used in this study are from East Asians. Kaplan-Meier curves were generated with UCSC Xena (<http://xena.ucsc.edu>) (Goldman et al., 2019).

Genetic variation data for the *ACE2* locus from 26 different populations around the globe were retrieved from the 1000 Genomes Project (The 1000 Genomes Project Consortium, 2015) at the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/>). The allele frequencies among East Asians were calculated from samples of the following populations: Chinese Dai in Xishuangbanna, Han Chinese in Beijing, Southern Han Chinese, Japanese in Tokyo, and Kinh in Ho Chi Minh City. The annotation for the genomic position (*e.g.*, protein-coding regions, introns, untranslated regions) and/or



consequences of variants (*e.g.*, synonymous, missense, stop-gained, etc.) was retrieved from the Ensembl Genome Browser (<https://www.ensembl.org/index.html>). The allele frequencies for *ACE2* in East Asians and in all populations were also retrieved from gnomAD (<https://gnomad.broadinstitute.org/>).

## **DECLARATIONS**

### **Acknowledgments**

We thank Drs. Di Liu (Wuhan Institute of Virology, CAS), Weiwei Zhai (Institute of Zoology, CAS), Yuting Zhao (UT Southwestern Medical Center), and Guanlin Wang (MRC WIMM Centre for Computational Biology, University of Oxford) for their critical comments and suggestions on the manuscript. The results shown in this study are in part based upon data generated by the TCGA Research Network (<https://www.cancer.gov/tcga>), the GTEx Project (<https://gtexportal.org/home/>), the 1000 Genomes Project (<https://www.internationalgenome.org/>), and the Genome Aggregation Database (<https://gnomad.broadinstitute.org/>).

### **Funding**

This work was done by the authors during their holiday break and was not supported by grants.

### **Authors' contributions**

Y.C., K.S., and W.Q. designed the study, performed data analyses, and wrote the manuscript.

### **Competing interests**

The authors declare that they have no competing interests.

### **Code availability**

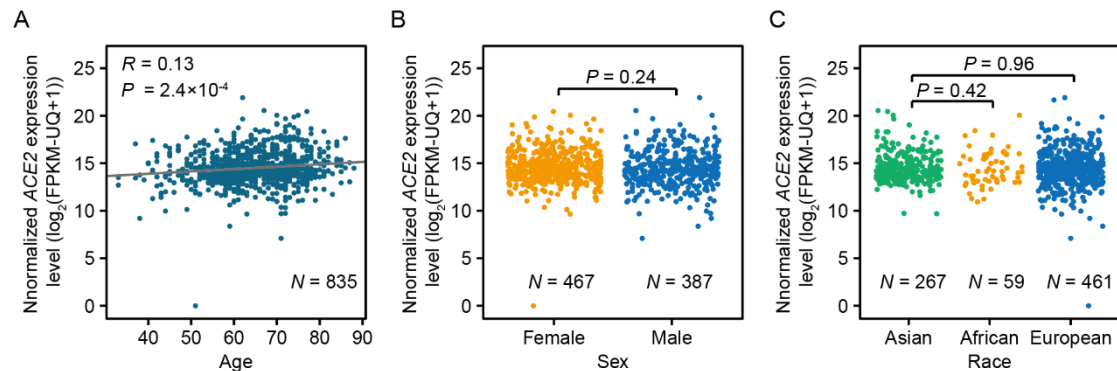
All scripts used to analyze the data and to generate the figures are available at <https://github.com/YingChen10/ACE2-Expression>.

## REFERENCES

- Chen, J., Yang, H., Teo, A.S.M., Amer, L.B., Sherbaf, F.G., Tan, C.Q., Alvarez, J.J.S., Lu, B., Lim, J.Q., Takano, A., *et al.* (2020a). Genomic landscape of lung adenocarcinoma in East Asians. *Nat Genet.*
- Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., Qiu, Y., Wang, J., Liu, Y., Wei, Y., *et al.* (2020b). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet.*
- Donoghue, M., Hsieh, F., Baronas, E., Godbout, K., Gosselin, M., Stagliano, N., Donovan, M., Woolf, B., Robison, K., Jeyaseelan, R., *et al.* (2000). A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circ Res* 87, E1-9.
- Fan, Y., Zhao, K., Shi, Z.L., and Zhou, P. (2019). Bat Coronaviruses in China. *Viruses* 11.
- Goldman, M., Craft, B., Hastie, M., Repčeka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., *et al.* (2019). The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv*, 326470.
- Gu, J., Gong, E., Zhang, B., Zheng, J., Gao, Z., Zhong, Y., Zou, W., Zhan, J., Wang, S., Xie, Z., *et al.* (2005). Multiple organ infection and the pathogenesis of SARS. *J Exp Med* 202, 415-424.
- Hamming, I., Timens, W., Bulthuis, M.L., Lely, A.T., Navis, G., and van Goor, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol* 203, 631-637.
- Hofmann, H., Geier, M., Marzi, A., Krumbiegel, M., Peipp, M., Fey, G.H., Gramberg, T., and Pohlmann, S. (2004). Susceptibility to SARS coronavirus S protein-driven infection correlates with expression of angiotensin converting enzyme 2 and infection can be blocked by soluble receptor. *Biochem Biophys Res Commun* 319, 1216-1221.
- Holshue, M.L., DeBolt, C., Lindquist, S., Lofy, K.H., Wiesman, J., Bruce, H., Spitters, C., Ericson, K., Wilkerson, S., Tural, A., *et al.* (2020). First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med.*
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., *et al.* (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet.*
- Jia, H.P., Look, D.C., Shi, L., Hickey, M., Pewe, L., Netland, J., Farzan, M., Wohlford-Lenane, C., Perlman, S., and McCray, P.B., Jr. (2005). ACE2 receptor expression and severe acute respiratory syndrome coronavirus infection depend on differentiation of human airway epithelia. *J Virol* 79, 14614-14621.
- Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., *et al.* (2019). Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*, 531210.
- Kuba, K., Imai, Y., Rao, S., Gao, H., Guo, F., Guan, B., Huan, Y., Yang, P., Zhang, Y., Deng, W., *et al.* (2005). A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus-induced lung injury. *Nat Med* 11, 875-879.

- Kuiken, T., Fouchier, R.A., Schutten, M., Rimmelzwaan, G.F., van Amerongen, G., van Riel, D., Laman, J.D., de Jong, T., van Doornum, G., Lim, W., *et al.* (2003). Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet* 362, 263-270.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., *et al.* (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*.
- Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., *et al.* (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426, 450-454.
- Lin, C., Song, H., Huang, C., Yao, E., Gacayan, R., Xu, S.M., and Chuang, P.T. (2012). Alveolar type II cells possess the capability of initiating lung tumor development. *PLoS One* 7, e53817.
- Mele, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D., Sullivan, T.J., *et al.* (2015). Human genomics. The human transcriptome across tissues and individuals. *Science* 348, 660-665.
- The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. *Nature* 526, 68-74.
- The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China. *Chinese Journal of Epidemiology* 41, 145-151.
- Uhlen, M., Fagerberg, L., Hallstrom, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Sjostedt, E., Asplund, A., *et al.* (2015). Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419.
- Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science*.
- Xu, X., Chen, P., Wang, J., Feng, J., Zhou, H., Li, X., Zhong, W., and Hao, P. (2020). Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci*, 1674-7305.
- Xu, X., Rock, J.R., Lu, Y., Futtner, C., Schwab, B., Guinney, J., Hogan, B.L., and Onaitis, M.W. (2012). Evidence for type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma. *Proc Natl Acad Sci U S A* 109, 4910-4915.
- Yan, R., Zhang, Y., Li, Y., Xia, L., and Zhou, Q. (2020). Structure of dimeric full-length human ACE2 in complex with B<sup>0</sup>AT1. *bioRxiv*, 2020.2002.2017.951848.
- Zhao, Y., Zhao, Z., Wang, Y., Zhou, Y., Ma, Y., and Zuo, W. (2020). Single-cell RNA expression profiling of ACE2, the putative receptor of Wuhan 2019-nCoV. *bioRxiv*, 2020.2001.2026.919985.
- Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., *et al.* (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*.

## FIGURES

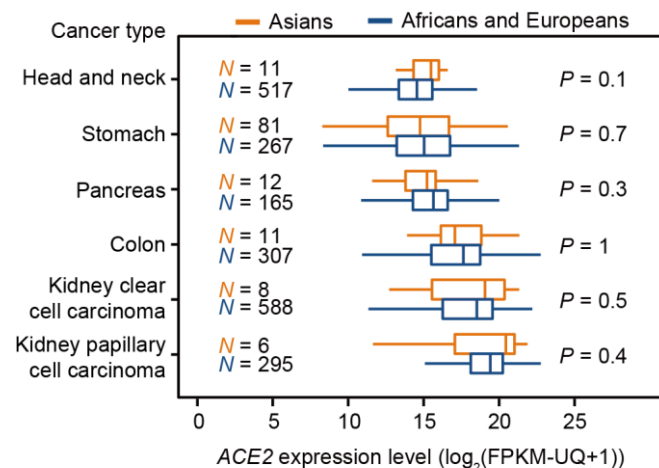


**Figure 1. *ACE2* expression levels are associated with age but not with sex or race, among LUAD samples.**

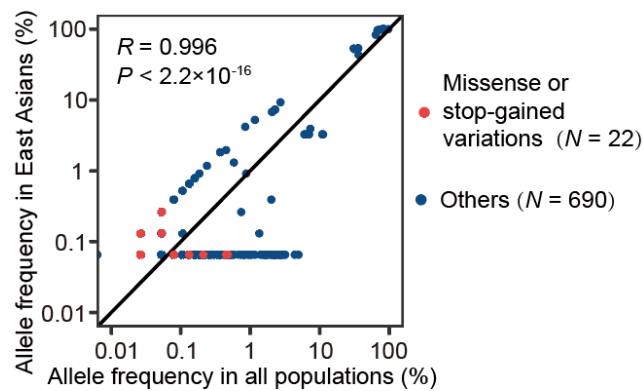
(A) Relationship between age and *ACE2* expression level. Pearson's correlation coefficient ( $R$ ) and the corresponding  $P$ -value are shown.

(B) Comparison of *ACE2* expression between sexes.  $P$ -value was calculated with a two-tailed Mann-Whitney  $U$  test.

(C) Comparison of *ACE2* expression among races.  $P$ -values were calculated with two-tailed Mann-Whitney  $U$  tests. The single sample of American Indian or Alaska Native in TCGA was not shown.



**Figure 2. Comparison of *ACE2* expression levels between Asians and others (Africans and Europeans) in additional tissues.** TCGA data for six cancer types in which the average *ACE2* expression level is higher than in LUAD samples are shown. Rectum adenocarcinoma also has *ACE2* expression higher than LUAD but is not shown, as there is only one available Asian sample.  $P$ -values were calculated with the two-tailed Mann-Whitney  $U$  tests. Outliers are not shown for the boxplots.



**Figure 3. Allele frequencies of *ACE2* variants detected by the 1000 Genomes Project, for East Asians ( $y$ -axis) or all populations ( $x$ -axis).** Each dot represents a polymorphic site, and the frequency of the alternate allele is shown. Only dimorphic genetic variations are shown. To avoid negative infinity during the logarithm transformation, we calculate the frequencies of alternate alleles that were not detected in East Asian samples as if 0.5 alleles were found. Detailed information is provided in Table S3.